



Bulletin de méthodologie sociologique

Bulletin of sociological methodology

85 | 2005
January

De quoi parlent les sociologues réunis en congrès ?

Éléments de complémentarité entre une analyse lexicale ouverte et le cumul de variables fermées

Gérard Boudesseul



Édition électronique

URL : <http://journals.openedition.org/bms/1041>

ISSN : 2070-2779

Éditeur

Association internationale de méthodologie sociologique

Édition imprimée

Date de publication : 1 janvier 2005

Pagination : 68-84

ISSN : 0759-1063

Référence électronique

Gérard Boudesseul, « De quoi parlent les sociologues réunis en congrès ? », *Bulletin de méthodologie sociologique* [En ligne], 85 | 2005, mis en ligne le 07 juillet 2008, consulté le 30 avril 2019. URL : <http://journals.openedition.org/bms/1041>

Ce document a été généré automatiquement le 30 avril 2019.

© BMS

De quoi parlent les sociologues réunis en congrès ?

Éléments de complémentarité entre une analyse lexicale ouverte et le cumul de variables fermées

Gérard Boudesseul

- 1 Le premier congrès de l'Association Française de Sociologie a rassemblé 1.015 intervenants ayant proposé une communication dans un réseau thématique, outre la participation de conférenciers invités. 127 d'entre eux, soit un peu plus d'un sur dix ont profité de l'occasion pour présenter deux à quatre communications, alors que quelques-uns plus rares, ont présenté une communication groupée, sur un total de 1.168 communications.¹ Ce congrès est l'occasion de donner une image synthétique des sujets de conversation qui animent les sociologues en France, que nous tenterons d'approcher au travers d'un recensement du lexique mobilisé. Deux approches sont ici confrontées : un regroupement par co-occurrences basé sur la fréquence d'emploi des mots deux à deux et une analyse plus discriminante fondée sur le calcul des écarts à la moyenne d'utilisation d'un mot.² Au préalable une brève présentation de la population des congressistes est nécessaire.

Qui sont les sociologues de l'AFS ?

- 2 Globalement, les fiches d'inscription fournissent quatre types d'indications : nom et prénom, l'origine géographique, l'origine institutionnelle, et le réseau de présentation. Nous avons réduit ces informations à quatre indicateurs : le sexe, l'origine parisienne, provinciale ou étrangère, l'origine institutionnelle réduite à l'annonce d'un laboratoire sans ou avant la mention d'une université, et enfin le numéro de réseau. L'assemblée des sociologues présente de grands équilibres : la répartition par sexe est globalement proche de la parité (H/F : 53/47%), cette proportion étant inversée lorsque l'origine est parisienne, en incluant les universités d'Ile-de-France. En second lieu, l'origine géographique se partage entre Paris-Ile-de-France et la Province (50/42), outre soixante-

dix collègues se localisant à l'étranger (6,3%), ce qui n'est pas négligeable du point de vue de l'écho d'une assemblée à ambition hexagonale. Enfin, la présentation au nom d'un laboratoire est très fréquente, les deux tiers des cas : elle peut être interprétée avec prudence, comme donnant le signe d'une insertion dans un laboratoire quel que soit le statut, qui n'était pas demandé, ou au moins d'une familiarité, à l'occasion par exemple, de séminaires de troisième cycle, pour les doctorants ou les jeunes docteurs. Il reste que plus d'un quart des intervenants mentionne une université mais pas de laboratoire. De plus, alors qu'à Paris 72,5% déclarent un laboratoire, 65,9% le mentionnent en province : négligence au moment de l'inscription ou distance à l'égard des laboratoires de recherche ? Sous réserve d'une fragilité de l'indicateur, la distance à l'égard des laboratoires serait plus grande en province. Quelle que soit l'origine géographique, on peut penser que ces intervenants sont plus jeunes, sinon sur le plan démographique, au moins du point de vue de la trajectoire de recherche.

- 3 Concernant l'origine institutionnelle, la nouveauté vient de la cinquantaine de sociologues qui ne mentionnent ni laboratoire, ni université. Ils se partagent en trois, selon qu'ils déclarent une entreprise, un service de la fonction publique et territoriale, et des individus. Cette nouveauté, encore modeste, traduit la volonté d'ouverture affichée par les fondateurs de l'association.
- 4 Ces grands équilibres masquent, comme bien souvent, une somme de déséquilibres, notamment dès lors que l'on considère le réseau d'adoption. Même si on ne peut exclure les aléas d'un tour de piste de découverte pour certains, les profils des réseaux sont souvent typés : les quarante réseaux diffèrent d'abord par leur poids, puisque le nombre de communications annoncées allait de 3 à 48, dont un tiers plus de 30 et une grande majorité entre 15 et 20 communications. Pour certains, on peut penser à un aboutissement mettant en commun des réseaux déjà existants, et pour d'autres un point de départ, alors que quelques intitulés assez proches font état de la préservation de collaborations, voire de traditions déjà établies ailleurs, comme l'ont revendiqué quelques spécialistes du travail, au nom du pluralisme. Dans cette logique multiforme, chaque réseau pourrait vivre en fonction de ses propres rythmes et non être soumis à une norme quelconque qui associerait un réseau à un champ ou à une approche, ou encore une norme moyenne de participation, relativisant par avance d'éventuels efforts de concentration ou de rationalisation des réseaux.
- 5 Le clivage de genre est le plus marquant : les femmes sont sur-représentées dans cinq réseaux (plus de 60%, soient 13 points de plus que la moyenne), autour de la famille, de la vieillesse, du genre, mais aussi des migrations, de l'environnement et des parcours de vie³. Alors que les hommes sont sur-représentés dans huit réseaux (plus de 66% soient 13 points de plus que la moyenne), autour des thématiques de la théorie et des auteurs, du sujet, mais aussi des médias et du sport, et enfin de l'espace politique, des associations, des classes et de la gestion⁴.
- 6 Ce seuil des deux tiers n'est atteint chez les femmes que dans les trois premiers réseaux cités. Une certaine division sexuée entre réseaux peut en être déduite, mais aussi une concentration féminine sur un nombre plus restreint d'entre eux. Peut-on parler d'une division sexuelle de travail sociologique ? Voyons ce que les congressistes disent, ou plus exactement ce que leur lexique nous dit sur eux.

Que disent les sociologues ?

- 7 L'analyse lexicale se fonde sur un recensement d'occurrences de mots, syntagmes ou groupes de mots entre eux, sans tenir compte, dans un premier temps au moins, des variables qui sont susceptibles de soutenir ces occurrences. Il en résulte des familles de formes lexicales, dépourvues en elles-mêmes de signification, mais que l'on peut ensuite reconstituer en imputant une logique propre à un enchaînement de mots fréquents⁵. À un extrême, un lexique uniforme traduirait un langage commun à l'ensemble des sociologues, voire une unité dans le choix des thématiques privilégiées. À l'autre extrême, un lexique très cloisonné traduirait un ancrage profond dans des objets de recherche disjoints. Une limite connue de ce type d'approche est que la répétition d'un mot ne traduit pas nécessairement un attachement à son contenu, mais éventuellement une forte hostilité : dans les deux cas, on mobilise des catégories typées. Une autre limite, connue de tous les correcteurs de copies d'examen, est que la répétition obstinée d'un mot trahit bien souvent une confusion quant à son contenu.
- 8 Les familles de formes lexicales, ou classes recensées, sont au nombre de cinq, elles-mêmes subdivisées en trois ou quatre sous-familles. Dans la première classe, représentant 28% des unités de contexte du corpus, les formes les plus usitées, classées par ordre décroissant de khideux sont : *sociologie, science, art, objet, théorie, intellectuel, contemporain, champ, concept, littéraire, auteur, philosophie, connaissance, critique, penser*⁶.
- 9 En dehors de *sociologie*, relativement consensuelle, se profile un niveau discursif de la science constituée et d'objets fixés dans une production littéraire, philosophique mais aussi éventuellement artistique. Bourdieu, ainsi que Weber, figurent en bonne place dans cette classe, à la fois comme référence et comme objet.
- 10 Dans la seconde classe, plus importante (34% du corpus), l'univers discursif semble à distance du précédent : les catégories mobilisées, toujours classées par ordre décroissant de khideux, sont *femme, familial, enfant, famille, homme, parent, sexe, couple, féminin, âge, jeune, immigré, masculin, genre*. Le regroupement est sévère par le confinement qu'il semble constater autour des femmes, de la famille, de la démographie et du genre. On est moins dans le domaine de la production littéraire que dans celui des composantes de la population et des relations qui les affectent. On ne sort de la sphère domestique que par la référence à *l'immigration* et, un plus loin dans le classement, aux *inégalités* dans *l'emploi* et la *scolarité*.
- 11 Ces deux classes mobilisent à elles seules près des deux tiers du corpus. Les trois suivantes se partagent le dernier tiers. Statistiquement moins représentatives du point de vue des occurrences, elles n'en conservent pas moins une grande visibilité. La troisième classe (18%) est aisément identifiable par les références à *l'entreprise* et au *client*, suivies des catégories *gestion, service, technologie, consommation, information, technique, salarié, relation, rationnelle, rationalisation, disposition, organisationnel, management, prescription*. On se situe ici clairement sur les terres de la sociologie du travail, de l'économie, et des organisations, avec une inflexion sur les dimensions relationnelles et prescriptives de l'entreprise. La quatrième classe (12%) s'articule autour des politiques publiques et de ses réformes : les catégories les plus mobilisées sont *politique, publique, réforme, santé, droit, gouvernement, dépense*, mais aussi *européen, loi, local, état, préventif, administration, législatif*. Un peu plus loin, on voit que le débat sur les politiques publiques est articulé avec celui

du pouvoir, de la *légitimité*, et de la *distance* entre le *local*, le *territorial* et les échelons *nationaux* et *européens*.

- 12 Enfin la portion congrue revient à la dernière classe (8%), elle aussi identifiable par ses formes les plus fréquentes : *méthode*, *analyse*, *approche*, *qualitatif*, *méthodologie*, *entretien*, *corpus*, *quantitative*, *donnée*, *logiciel*, *compléter*, *enquête*, *questionnaire*, *empirique*, *observation*. *Méthode* et *méthodologie* concentrent le vocabulaire de cette classe autour du recueil et du traitement des matériaux, ainsi que dans des postures formulées de manière ouverte en termes d'*approche*, *observation* et du verbe *compléter*.
- 13 Ce découpage paraît banal si on réduit les familles à des libellés un peu trop standardisés⁷. Il l'est moins si on détaille la cohérence interne de chaque classe.

Une théorisation à géométrie variable

- 14 La classe 1 *auteurs-théorie*, se décompose en trois sous-familles. La première est centrée sur le *champ*, notamment *littéraire*, la *philosophie* et les *auteurs*. La référence à *l'écriture* suggère une approche de la *société* au travers des œuvres. Pourtant, quelques objets très contemporains donnent l'occasion d'emprunter à ce lexique : les *médias* et le *sport*, avec un degré d'association modeste, il est vrai, au regard des autres formes (khideux respectifs de 33 et 26).
- 15 La deuxième sous-famille fait plutôt grand cas de la *science* et de la *théorie*, de la *connaissance* et du *sujet*, en s'appuyant sur des références à *l'histoire* et à *Weber*. L'approche est moins attachée aux œuvres qu'aux *paradigmes*, et aux registres *symbolique* et *idéologique*. L'objet *environnement* s'y est glissé avec un degré d'association plus faible que les précédents (khideux de 15).
- 16 Enfin, la troisième, un peu intermédiaire, associe l'*objet* et le *concept* avec l'*art*, le *contemporain* et... *Bourdieu*, mais aussi les *prisons* et le *militaire* ! D'un côté, le registre est conceptualisé et emprunte volontiers à un auteur devenu canonique, de l'autre il est particulièrement, mais non exclusivement, mobilisé sur des objets spécifiques, tels que la production artistique, la musique mais aussi les prisons et l'armée.

Le cheminement du genre aux inégalités

- 17 La classe 2 *société-inégalités* semble confirmer la concentration des catégories du genre autour de l'étude de la sphère domestique et de la population. La première sous-famille cumule en effet quelques-unes des catégories phare de cette classe : *famille* et *familial*, *enfant*, *parent*, *couple*, *âge*, *jeune*, mais aussi *homosexualité*. La seconde, proche, mêle *genre* et *migration* associée à ses formes dérivées : *immigr(er,é)*, *migrant*, *migratoire*. Elle mentionne ensuite la *stigmatisation*, la *ségrégation* et les *minorités*. Ces deux sous-familles demeurent dans un périmètre qui relèverait de la construction sociale des catégories démographiques de la population. Toutefois, les deux autres sous-familles débordent nettement de ce périmètre. L'une, polarisée par les catégories *femme*, puis *homme* et *sexe*, outre leurs adjectifs respectifs, mentionne l'*emploi* avant le *domestique*, ainsi que la notion d'*inégalité*. L'autre se réfère à la *génération*, mais cumule ensuite *origine*, *parcours*, *carrière* et *trajectoire* avec des références de lieu tels que l'*école* et ses formes dérivées, *scolaire* et *scolarité*, ainsi que le *quartier*, termes associés aussi aux catégories classiques de la tradition sociologie : *classe* et *populaire*. Finalement, le passage en revue détaillé en dit plus sur le profil global de cette classe que les formes statistiquement les plus fréquentes

ne le laissent penser. Les catégories en terme de genre se sont diffusées dans l'analyse des migrations, de l'emploi, de l'école et des classes sociales qui n'apparaissent ensuite dans aucune autre classe lexicale en tant que formes significatives.

Une économie relationnelle

- 18 La classe 3 *travail-entreprise* semble opposer une dimension relationnelle à une dimension gestionnaire du management à base de technologie et de prescription. La première est illustrée par deux sous-familles. L'une associe étroitement les adjectifs *relationnel*, *marché*, *marchand* à *consommation*, mais aussi à *collectif* et *productif*. L'autre associe la *relation*, *service*, *client* avec le *travail*, *l'entreprise*, le *salarié*, la *compétence* et *l'organisation*. Cette dimension relationnelle mêle donc les relations dans l'entreprise et avec le consommateur. La dimension gestionnaire réunit également deux sous-familles. La première est polarisée par la *gestion*, mais aussi la *technologie* et la *rationalisation*, alors que la seconde mentionne d'abord le *management*, la *prescription* et le *contrôle*, ainsi que *l'informatique*. Une autre opposition apparaît en filigrane : alors que la dimension relationnelle est développée dans les *services*, la dimension technique et gestionnaire est associée à *l'industrie*.

De la « policy » au « politics », bivalence ou ambivalence ?

- 19 La classe 4 *politiques publiques* comporte tout d'abord deux sous-familles nettement orientées vers la *réforme* ou la *régulation* de certaines *politiques publiques* accompagnées d'une référence explicite au *gouvernement*, à *l'Etat* et au *pouvoir*. La seconde pourrait tenir lieu d'illustration avec une mention prioritaire aux *dépenses de santé*, à la législation, mais aussi à des méthodes politiques : *prévention*, *proximité*, *participation*. Les deux autres sont plus éloignées de cette perspective : l'une est centrée sur le *droit* et les *lois* avec une dimension *européenne*, mais aussi *locale*, *régionale*, *nationale* et enfin, *associative*. L'autre se rapproche plus nettement du Politique que des politiques gouvernementales en évoquant la question de la *légitimité*, de *l'institution*, du *conflit* et de la *négociation*. Dans l'ensemble, la classe reste plutôt orientée vers les politiques sociales et les enjeux de leur réforme.

Méthode et discours de la méthode

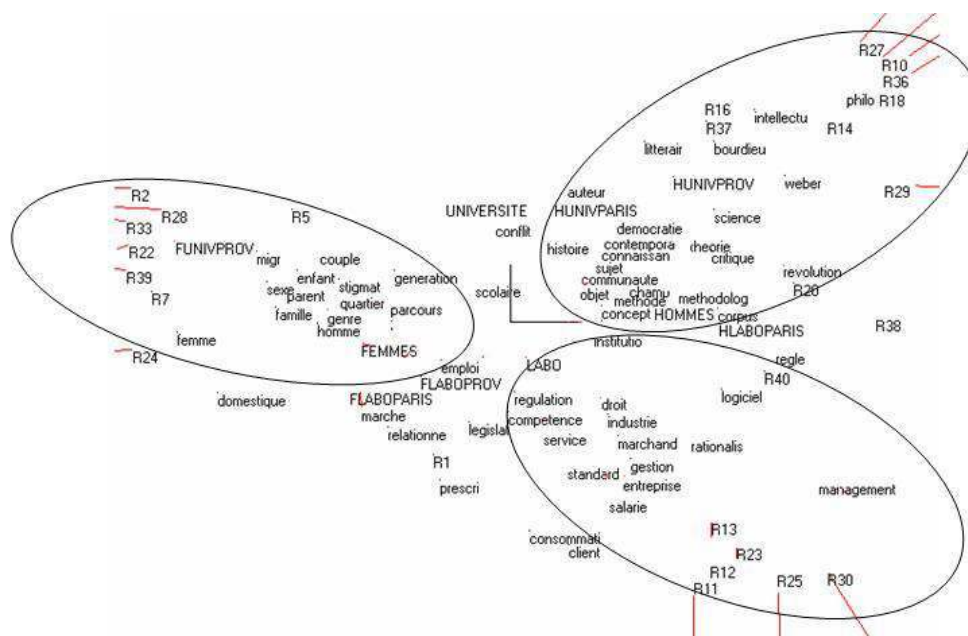
- 20 La classe 5 *méthode* est curieusement isolée, alors qu'on pourrait s'attendre à ce qu'elle imprègne l'ensemble des communications. *Méthode* et *méthodologie* sont les plus caractéristiques de la première sous-famille avec *données*, *recueil*, *approche* et *analyse*, dans une *dimension empirique*, que l'on trouve associée de manière inattendue à un objet, l'*électorat* (et *électoral*). A proximité, la seconde sous-famille unit un vocabulaire de *corpus*, *enquête*, *questionnaire*, *typologie*, avec une référence à *l'interactionnisme*. Dans le même ordre d'idée, *observation*, *entretien*, *qualitatif* et *quantitatif* sont associés à la *complémentarité*, mais aussi à *l'anthropologie*. Enfin, *série*, *échelle* et *territoire* sont associés à *démarche*, *comparaison* et *comparatif*. Au total, ces quatre sous-familles sont proches les unes des autres et leur examen ne modifie guère le profil global de la classe.
- 21 En somme, le fait que deux tiers du corpus soient rassemblés dans deux classes peut-être pris comme une sorte de consensus sur les objets : à côté des sociologies spéciales du politique, du travail et de la méthode, une culture commune rassemblerait des

sociologues de tous bords autour d'une étude du lien social d'un côté, et d'une tradition sociologie incarnée par ses oeuvres de l'autre. Or, la géographie détaillée des sous-familles montre des déplacements de champs et d'approches : dans la classe 1, théorie, paradigmes et sujet étaient certes l'objet principal de quelques réseaux, mais, ces catégories semblent plus spécifiquement revendiquées dans l'étude de certains objets contemporains comme l'art, les médias, l'environnement, le sport, et le militaire. Une telle spécificité reste à expliquer. Dans la classe 2, la question traditionnelle des inégalités face à l'école et à l'emploi semble avoir été lue en priorité en fonction de l'âge et du sexe, ce qui constitue en quelque sorte un retour aux variables fondamentales de la population, minorant du même coup les références historiques à la reproduction. Toutefois, les concepts plus construits de genre et de génération semblent devenir des outils en voie de diffusion. *A contrario*, les catégories associées au genre restent aussi très ancrées dans l'analyse de la sphère domestique, de la conjugalité et de la filiation. Pour les trois autres classes, on ne sait si leur faible part dans le corpus doit amener à les considérer comme minoritaires, ou si, justement, elles sont un peu utilisées par tous les congressistes. On touche ici à une limite du recensement des occurrences, qui invite à se retourner vers la statistique des auteurs de communications. En particulier, ce recensement ne permet pas de savoir jusqu'à quel point les variables attachées aux personnes sont entièrement ou partiellement déterminantes du vocabulaire reconstitué en unité de contexte.

Qui dit quoi ?

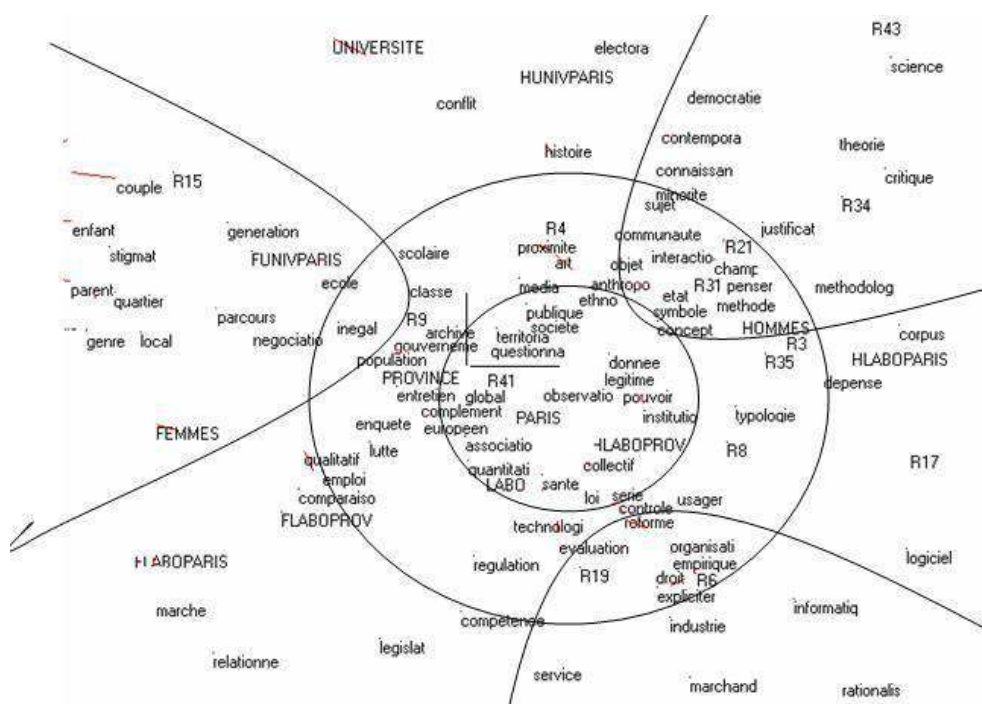
- 22 Une autre méthode consiste à représenter graphiquement l'analyse factorielle des occurrences, en incluant des variables ayant un potentiel explicatif. Cette représentation suggère un paysage un peu différent de la précédente, avec trois classes distinctes du point de vue de leurs contributions aux facteurs⁸. Ces trois classes sont la classe *société-inégalités*, la classe *travail-entreprise* et la classe *théorie-auteurs*, suggérant une représentation en forme d'hélice d'avion à trois pales. Considérés comme variables actives, une vingtaine de réseaux se situent aux extrémités de ces pales (voir graphique 1).

Graphique 1 : Analyse factorielle des occurrences



- 23 Chacune des trois classes se répartit autour de quelques mots dont la contribution aux facteurs est la plus élevée. Il n'est pas certain qu'il faille absolument chercher à nommer les axes, ce qui réduirait la logique propre de cette représentation hélicoïdale : chaque classe a sa cohérence et s'oppose aux deux autres, comme l'a déjà montré l'analyse lexicale. C'est pourquoi on serait tenté de parler d'attraction autour de quelques formes lexicales ou de variables actives plutôt que d'attraction le long d'un axe, rapprochant ainsi l'analyse lexicale qui procède par repérage linéaire et par arborescence des co-occurrences, et la représentation topographique multidimensionnelle de l'analyse factorielle (en fait, le plus souvent bi ou tri-dimensionnelle).
- 24 Ensuite, certaines formes lexicales se présentent comme une charnière entre classes :
- les mots *emploi*, *marché*, *relationnel* sont ainsi à mi-chemin des classes *société-inegalités* et *travail-entreprise*, suggérant deux rapprochements possible : soit la dimension relationnelle qui relèverait plutôt de la classe *société-inegalités* gagnerait la sphère de l'économie, soit l'accès à l'emploi et de plus en plus questionné pour des catégories de population selon l'âge, le sexe et la nationalité.
 - les mots *scolaire*, *classe* et *génération* se présentent eux comme une charnière entre les classes *société-inegalités* et *théorie-auteurs* : on peut supposer que les analyses en terme d'âge et de sexe, vus comme variables empiriques, ont contribué à détacher les champs de l'école et des classes sociales du registre théorique dans lequel on a souvent dit qu'elles étaient restées confinées.
 - enfin, les mots *institution* et *règle* se présentent comme limitrophes de la classe *travail-entreprise* et *théorie-auteurs*. On peut supposer qu'ils révèlent une lecture plus institutionnaliste et conventionnelle de l'économie de marché, en même temps qu'ils témoignent de la recherche d'un substrat empirique au concept canonique de règle, lequel emprunte aussi par ailleurs à la méthodologie dans l'histoire de la discipline.
- 25 Le reste des formes lexicales est très concentré aux alentours de l'origine des axes avec une faible contribution aux facteurs de l'analyse. Doit-on en conclure à une faible signification ou à une banalisation ? Si l'on focalise l'observation sur ce qui semble un

Graphique 2 : Zone de convergence entre trois classes



- 26 Chacune des trois classes les plus typées semble emprunter aux lexiques politique et méthodologique. Cette représentation, plus concentrique qu'hélicoïdale, suggère un autre consensus : celui d'un accent fréquent sur la dimension politique des phénomènes observés, et de la justification des méthodes employées. Moins les points sont dispersés sur le graphique, plus ils appartiendraient à la fois aux différentes classes.
- 27 L'introduction de variables permettant de préciser le profil des congressistes au regard du répertoire qu'ils mobilisent amène à douter de cette rassurante convergence. La solution la plus simple consiste à les joindre à l'analyse factorielle au titre de variables « supplémentaires » : on trouve ainsi confirmation d'une localisation plutôt distante entre hommes et femmes, mais les positions des variables de lieu (Paris, province) et de cadre (laboratoire, université) n'est guère probante (cf. graphique 1). Plus encore, le proximité d'une de ces variables à l'égard d'un axe peut laisser penser que cette proximité est vraie pour tout individu présentant la variable et que celui-ci est toujours associé de manière équivalente aux points qui l'entourent.
- 28 C'est à ce point que la quantification de l'attraction entre variables et formes lexicales est utile ; elle permet de repérer dans quelle mesure précise elles sont associées. Le calcul du « pourcentage d'écart à la moyenne » (PEM) de chaque variable et de chaque forme lexicale autorise cette confrontation. Dans cette phase, on opère un nouveau va-et-vient entre l'approche lexicale de libellés ouverts, et l'approche des PEM sur des questions fermées ⁹. A été extraite du corpus une liste de 114 mots-clés, caractéristiques de chaque classe et sous-famille selon Alceste, pour lesquels on a ensuite procédé aux calculs de PEM. On peut ensuite ranger ces mots-clés dans les classes dessinées par l'analyse

lexicale. On obtient ainsi un tableau associant chaque variable dessinant le profil des congressistes, aux mots les plus employés¹⁰.

- 29 Il en ressort que la préférence lexicale des hommes va à la classe *théorie-auteurs* pour un tiers des mots qu'ils utilisent, et que la préférence des femmes va à la classe *société-inégalités* pour 56% des mots qu'elles utilisent. Une préférence de second ordre est commune aux deux sexes : un quart des mots-clés appartiennent à la classe *travail-entreprise*. Les variables de cadres interviennent dans ce classement, puisque la référence à un laboratoire est pour beaucoup dans ce registre commun : 38% des mots utilisés par les laborant(e)s¹¹, favorisent dans une proportion proche la classe *méthode*. Une hypothèse plausible serait qu'un travail collectif encadré par un laboratoire favorise une communauté d'objet et de langage. En revanche, l'origine géographique aurait un effet discriminant : les sociologues parisien(ne)s expriment une préférence lexicale marquée pour la classe *politiques publiques*, alors qu'en province, comme chez les femmes en général, la classe *société-inégalités* vient en premier. Que l'origine soit parisienne ou provinciale, la classe *travail-entreprise* vient aussitôt après. En résumé, le laboratoire semble réunir alors que l'origine géographique discriminerait plutôt le répertoire des sociologues, sauf pour la classe *travail-entreprise*. Ces hypothèses restent fragiles dans la mesure où ces variables agissent simultanément.
- 30 Un moyen de dissocier les variables pour identifier des effets propres, sans pour autant recourir à des modèles complexes, consiste à isoler quelques profils de congressistes, ce que permet Trideux. Les combinaisons des trois variables dichotomiques de sexe, lieu et cadre aboutissent à huit profils-types. On découvre alors d'autres effets qu'il convient de distinguer, selon ce qu'il advient du lexique commun, des clivages déjà constatés, ou de nouvelles formes de clivages.
- 31 Tout d'abord, l'existence d'un lexique commun à la classe *travail-entreprise* constatée selon les variables homme, femme et laboratoire prises séparément, est confirmé lorsque ces variables sont cumulées. Les laborantines provinciales et les laborantins provinciaux lui consacrent la plus grande part de leur répertoire (lecture horizontale du tableau), et c'est chez eux que l'on constate le plus grand intérêt pour cette classe (lecture verticale). Dans le même ordre d'idées, les laborantines parisiennes et les laborantins parisiens lui consacrent aussi une part importante de leur répertoire, avec en complément, une importance spécifique accordée à la classe *méthode*. On pourrait légitimement conclure à un effet propre du laboratoire avec une accentuation méthodologique dans les laboratoires parisiens.
- 32 Qu'advient-il des clivages constatés dans le cas général, suggérant un attrait particulier des femmes pour la classe *société-inégalités* et des hommes pour la classe *théorie-auteurs* ? Le choix de la classe *société-inégalités* apparaît très renforcé chez les provinciales universitaires (69%), au point d'écraser les autres classes, de même que pour les parisiennes universitaires, à un moindre degré (38%). On peut se demander si l'absence de mention d'un laboratoire ne signifie pas un faible encadrement, qui exposerait plus les étudiantes au sens commun de ce qu'une jeune femme peut choisir comme objet de recherche. Des entretiens approfondis sur les critères de choix des sujets de mémoires en Master 1 et 2 seraient ici sans doute éclairants. Concernant le choix de la classe *théorie-auteurs* par les hommes, il se trouve renforcé par les provinciaux universitaires (44%) et ensuite par les parisiens universitaires (31%) qui trouvent ici leur classe favorite. On peut se demander si la référence aux auteurs n'est pas considérée comme un attribut de sociologues masculins lorsqu'ils se définissent comme universitaires, c'est-à-dire sans

doute fraîchement diplômés et socialisés à la recherche, sans pour autant être engagés dans un laboratoire (répétons que les « universitaires » incluent ici les étudiants). La distance à l'égard d'un laboratoire ne prive donc pas de la référence aux auteurs, comme objet ou comme méthode. Des entretiens approfondis sur la construction de l'image de ce qu'est la sociologie auprès des étudiants de licence nous permettrait d'en savoir plus. Ces deux sources de clivage nous ramène finalement à la répartition sexuée des congressistes entre les réseaux et à l'hypothèse d'une division sexuelle du travail sociologique.

- 33 A ce stade, les données permettent d'associer les classes lexicales aux variables dessinant le profil des congressistes. Plus le nombre de variables disponibles est élevé lors du recueil de données, plus on sera en mesure d'affiner le profil des personnes. D'un côté ceci permettra de dresser des profils types de plus en plus précis, d'un autre côté, l'effet propre de chaque variable sera mieux cerné. En se rapprochant de la dimension individuelle, on est alors mieux fondé à retourner au corpus des communications et choisir des illustrations, ou mieux, procéder à des entretiens approfondis, ce que l'espace de cet article ne nous permet pas de faire.

Qui dit quoi selon le contexte de la prise de parole

- 34 Un dernier exercice, un peu plus fouillé, est envisageable si l'on essaie de déconstruire la variable sexe, qui paraît ici déterminante, et dont on pourrait croire, à tort, qu'elle est exogène : le discours des hommes est-il toujours conforme au profil masculin, et celui des femmes au profil féminin ? Si le croisement de variables a montré qu'il n'en va pas toujours ainsi dans le cadre d'un laboratoire, on peut aussi le vérifier en s'interrogeant sur l'effet propre de la mixité des réseaux : ai-je tendance à m'exprimer différemment devant une assemblée dans laquelle l'un des deux sexes domine statistiquement ?
- 35 Si on isole de part et d'autre les réseaux à faible mixité, on peut reproduire le même type de raisonnement que précédemment et tester les profils types de congressistes. Compte tenu de la concentration des femmes dans un nombre plus restreint de réseaux, nous avons cherché un critère qui permette d'équilibrer deux sous-corpus de taille équivalente. Ce critère est satisfait lorsqu'on se limite aux réseaux présentant un sex-ratio déséquilibré et, de plus, renforcé par un déséquilibre dans l'une ou l'autre des variables disponibles, cadre institutionnel ou lieu. On arrive ainsi à la répartition suivante, dans laquelle il ressort que le déséquilibre de sex-ratio se cumule souvent à un cadre universitaire (plus que de labo) chez les hommes, et à une origine parisienne chez les femmes.

Tableau 1 : Réseaux présentant un profil renforcé d'intervenants (unité : numéro de réseau)

	Réseaux à dominante masculine			Réseaux à dominante féminine		
	Paris	Province	Lieu partagé	Paris	Province	Lieu partagé
Laboratoire	35			11	38	22
Université	27, 37	31	36	15	7	
Cadre partagé	34	5		24,33,42		

Lecture du tableau 1 : Les congressistes du réseau 35 sont à dominante masculine, parisienne et déclarent le plus souvent un laboratoire

- 36 Le corpus des communications dans ces réseaux étant isolé du reste, on peut constituer un tableau du lexique le plus fréquent pour chaque profil, comme précédemment (cf. en annexe le tableau 5). Dans les réseaux à dominante masculine, tout d'abord, les hommes accentuent nettement la préférence pour la classe *théorie-auteurs* (44%), déjà dominante dans le cas général, suivie de la classe *méthode*, au détriment de la classe *travail-entreprise*, alors que les femmes maintiennent en priorité les deux classes mentionnées pour le cas général, à savoir les classes *travail-entreprise* et, à un moindre degré, *société-inégalités*. Mais dans le détail, un fait nouveau apparaît : la préférence pour la classe *théorie-auteur* n'est plus exclusivement masculine. Les femmes laborantines et parisiennes opèrent un choix analogue à celui des hommes dans ce type de réseaux, avec un accent très marqué pour la classe *théorie* (50%), suivi de la classe *méthode*. Les laborantines provinciales, elles, rejoignent le choix *société-inégalités* déjà constaté dans le cas général.
- 37 Dans les réseaux à dominante féminine, les hommes s'éloignent du cas général et retrouvent plutôt le choix des laborantins parisiens avec une prédominance du lexique commun *travail-entreprise*. De leur côté, les femmes endossent dans ces réseaux des attributs dont on a vu dans les réseaux masculins qu'ils pouvaient être partagés : la référence à la classe *théorie-auteurs* y est pour elles la plus élevée (31%) sans pour autant sacrifier la classe *travail-entreprise* commune. Si l'on procède au cumul des variables afin d'affiner encore les profils, on constate que chez les hommes, la référence à un lexique commun *travail-entreprise* est spécifique aux laboratoires provinciaux (44%), alors que chez les femmes la préférence pour la classe *théorie-auteurs* et la classe *travail-entreprise* est plus fréquente chez les parisiennes que chez les provinciales, qu'elle revendiquent ou non un laboratoire. Pour les autres profils, on retrouve le cas général.
- 38 En résumé, de curieux effets de renforcement semblent se produire :
- hommes et femmes mobilisent nettement plus le registre *théorique* lorsqu'ils sont dans un réseau dont le sex-ratio leur est favorable.
 - les femmes d'origine parisienne ont la spécificité d'adopter ce registre théorique dans les deux types de réseaux à sex-ratio déséquilibré, surtout lorsqu'elles mentionnent un *laboratoire*. Faut-il voir là une expression de la légitimité conférée par l'appartenance à un laboratoire parisien, devant une assemblée qui peut être très masculine, mais aussi très féminisée ? On pourrait opposer à ceci que les hommes d'origine parisienne adoptent aussi ce profil quel que soit le réseau selon son degré de mixité. Mais pour eux, il est vrai, il ne s'agit que de retrouver le profil masculin général. Ce résultat ne résulte pas pour eux de l'effet propre d'une faible mixité.
 - l'existence d'un *lexique commun autour du travail et de l'entreprise* est confirmé dans la plupart des profils, à l'exception des hommes dans les réseaux où ils sont majoritaires, et lorsqu'ils ne mentionnent pas de *laboratoire*. Cette communauté d'objets et de discours est particulièrement visible chez les hommes insérés dans des réseaux féminisés, et parmi ceux-ci dans les laboratoires provinciaux. A y regarder de près, c'est aussi vrai dans les laboratoires provinciaux de réseaux masculinisés (31%).
 - enfin, on peut, au bas mot, parler d'*inertie* pour les profils particuliers qui confirment la tendance générale, quel que soit le type de réseaux, mixte ou non ? La classe *société-inégalités* est significativement associée aux femmes, particulièrement dans les universités pour le cas général comme dans les réseaux fortement féminisés. C'est aussi vrai auprès des femmes provinciales dans les réseaux à dominante masculine. Symétriquement, la classe *politiques publiques* spécifique aux universités parisiennes et plutôt appropriée par les hommes dans le cas général, est partagée avec les femmes dans les réseaux masculins à condition de se

trouver dans ces mêmes universités parisiennes. Les femmes de laboratoire provinciaux ne trouvent ce profil que lorsqu'elles sont dans des réseaux féminisés. Mais dans l'ensemble, la part du lexique relevant de cette classe *politiques publiques* est deux fois plus élevée chez les hommes que chez les femmes, dans le cas général, comme dans les réseaux féminisés.

- 39 Ainsi, selon la méthode employée, les limites d'un consensus lexical se déplacent et avec elles les espaces d'une communauté d'objets et de méthodes entre les sociologues (cf. en annexe le tableau 10).
- 40 Si l'on prend les mots comme seules variables, les deux méthodes illustrées par les deux logiciels suggèrent une visualisation différente : l'analyse lexicale (Alceste) met l'accent sur la fréquence des co-occurrences ; c'est-à-dire, sur ce qui rapproche les mots. Les cas les plus fréquents se rencontrent donc dans la classe *société-inegalités* et *théorie-auteurs*. Le calcul de pourcentage d'écart à la moyenne (Trideux) met l'accent sur les écarts, et ce qui met en position comparable des écarts semblables. Dans ce cas, ceux qui s'écartent le moins de la moyenne se retrouvent dans les classes *méthode* et *politiques publiques*.
- 41 Si l'on prend maintenant comme variables les attributs des auteurs associés aux mots employés, plus on affine la focale, plus l'hétérogénéité s'accroît, un peu comme dans les enquêtes de mobilité. On peut déduire de cette hétérogénéité quelques effets discriminants, mais aussi les classes les plus consensuelles. On combine alors les apports des deux méthodes en conservant les familles de mots, ainsi que ceux d'entre eux qui sont les plus significatifs, pour ensuite les croiser avec les attributs des individus. L'usage de ces attributs peut être progressif : usage *séparé* d'abord, pour évaluer l'effet brut d'une variable, usage *cumulé* ensuite, permettant de dessiner des profils d'individus associés à un lexique typé, usage *contextualisé*, enfin, lorsque ces profils sont re-situés dans des sous-ensembles du corpus : il faut alors retourner au corpus initial et le décomposer selon des critères devenus apparemment décisifs : dans le cas présent, les réseaux (variable de contexte 1), et ceux dont le sex-ratio est nettement déséquilibré (variable de contexte 2).
- 42 Au fil des trois usages possibles des variables, la classe « consensuelle » ; c'est-à-dire, celle qui donne le moins lieu à des effets discriminants, demeure la classe *travail-entreprise*. Or, elle n'est mentionnée en ce sens par aucune des deux méthodes reposant uniquement sur les mots, privés de variables (familles lexicales pour Alceste, calcul de PEM pour Trideux). A l'opposé, la classe *société-inegalités* s'avère la plus discriminante, quelque soit l'usage des variables, ce que suggèrait déjà l'analyse factorielle. Enfin, on pouvait s'attendre à ce que le passage des variables séparées aux variables cumulées, puis contextualisées, fasse ressortir de plus en plus d'effets de discrimination. C'est vrai pour le cumul de variables qui affine les profils des individus, mais la prise en considération d'un effet contextuel montre aussi la délimitation d'espaces communs. Elle suggère des lieux et des circonstances par lesquels peuvent se construire des objets et des méthodes partagés. Dans ce cas, les méthodes d'enquêtes les plus proches du recueil de récits pourrait reprendre leurs droits : guidées par ce ciblage, elle seraient plus à même de saisir la genèse de ces cohérences, qui sont locales dans leur appropriation et leur manifestation, et à la fois locales et globales dans leurs conditions d'émergence et de signification.

BIBLIOGRAPHIE

Philipp Cibois (1984), Méthodes post-factorielles pour le dépouillement d'enquêtes, *Bulletin de Méthodologie Sociologique*, n. 1, octobre, pp. 41-78.

Philipp Cibois (1995), Trideux version 2.2, *Bulletin de Méthodologie Sociologique*, n. 46, mars, pp. 119-124.

Max Reinart (1987), Classification descendante hiérarchique et analyse lexicale par contexte: Application au corpus des poésies d'A. Rimbaud, *Bulletin de Méthodologie Sociologique*, n. 13, janvier, pp. 53-90.

Max Reinart (1990), Alceste - Une méthodologie d'analyse des données textuelles et une application: Aurélia de G. de Nerval, *Bulletin de Méthodologie Sociologique*, n. 26, mars, pp. 24-54.

NOTES

1. Nous n'avons compté qu'une fois dans la statistique les communications comportant le même titre et le même résumé dans le même réseau, sauf lorsque leurs auteurs étaient différents ou présentaient des attributs différents.
2. Les logiciels utilisés sont Alceste pour l'analyse lexicale et Trideux pour le calcul des écarts à la moyenne. Pour une présentation synthétique des deux logiciels, on consultera le site de l'Université de Versailles-Saint-Quentin ainsi que le B.M.S. n° 26 de mars 1989. Voir Cibois, 1984 & 1995; Reinart, 1987 & 1990.
3. Réseaux 33, 38, 7, 24, 2, 22, classés par ordre décroissant. Les thématiques données ici ne sont que suggestives afin d'alléger les nombreux renvois qui suivent. Elles ne sauraient condenser les intitulés complets et les problématiques, du reste évolutives, affichées par les fondateurs de chaque réseau.
4. Réseaux 26, 18, 37, 31, 34, 35, 5, 30.
5. Le logiciel utilisé est ici Alceste, qui fournit en outre des citations contextualisées par les « unités de contexte », ensemble de quelques lignes par lequel le logiciel découpe le corpus des communications.
6. Chaque substantif est ici étendu à son pluriel ainsi que son adjectif s'il a le même radical. Lorsque l'adjectif est cité, il n'inclut pas le substantif. Par exemple le mot *publique* n'inclut pas *public* qui couvre un champ sémantique beaucoup plus large, outre le fait qu'il est à la fois adjectif et substantif.
7. Comme nous le ferons nous mêmes *infra*, par commodité, pour renvoyer à ces classes autrement que par des numéros : classe 1 *théorie-auteurs*, classe 2 *société-inégalité*, classe 3 *travail-entreprise*, classe 4 *politiques publiques*, classe 5 *méthode*.
8. Les valeurs propres de chaque facteur expliquent respectivement 11,8%, 8,4% et 6,2% du phi-deux total.
9. Le fondement théorique de cette confrontation est celui de l'enrichissement de variables qualitatives ouvertes par leur croisement avec plusieurs questions fermées cumulées. Pour une justification et une illustration de l'usage du PEM, voir Cibois P., « Eclairer le vocabulaire des questions ouvertes par les questions fermées : le tableau lexical des questions », *Bulletin de Méthodologie Sociologique*, 26, 1989, pp. 12-23.

10. Tableau réduit aux 16 premiers mots, classés par ordre décroissant de PEM positif. Cf. en annexe, le tableau 4. Le lexique est rangé par classe dans le tableau 7. L'extraction d'un lexique réduit ici à 114 mots facilite l'usage du logiciel Trideux qui ne peut traiter un corpus aussi important que l'ensemble des communications du millier de congressistes.

11. Afin d'alléger la lecture, nous appellerons laborantins, universitaires, parisiens et provinciaux les congressistes ayant mentionné ces appartenances sur leur fiche d'inscription, sachant pertinemment qu'il ne s'agit que de déclarations transitoires, dans un contexte donné, et que de surcroît, la référence à un laboratoire s'accompagne sans doute dans la plupart des cas d'une inscription universitaire.

RÉSUMÉS

Le premier congrès de l'Association Française de Sociologie a fourni l'occasion de donner une image synthétique des sujets de conversation qui animent les sociologues en France, que nous tenterons d'approcher au travers d'un recensement du lexique mobilisé. L'article confronte ce type d'approche à une analyse plus quantifiée, fondée sur le calcul des écarts à la moyenne, pour enfin chercher comment les deux approches peuvent être combinées. L'analyse lexicale, à l'aide du logiciel Alceste, met l'accent sur la fréquence des co-occurrences, c'est-à-dire sur ce qui rapproche les mots. Les occurrences les plus fréquentes se rencontrent dans une classe *société-inegalités* et une classe *théorie-auteurs*. Le calcul de pourcentage d'écart à la moyenne (logiciel Trideux) souligne les écarts, et ce qui met en position comparable des écarts semblables. Dans ce cas, les mots qui s'écartent le moins de la moyenne d'utilisation et seraient les plus partagés par l'ensemble des sociologues, se retrouvent dans les classes *méthode* et *politiques publiques*. Si l'on introduit comme variables les auteurs, avec pour attributs, le sexe, l'institution d'origine, le lieu, et qu'on les associe aux mots qu'ils emploient, on combine alors les apports des deux méthodes en conservant les familles de mots ainsi que ceux d'entre eux qui sont les plus significatifs, pour ensuite les croiser avec les attributs cumulés des individus. La variable la plus discriminante est alors le sexe, avec une concentration des femmes dans la famille *société-inegalités* et une concentration des hommes dans la famille *théorie-auteurs*. La classe la plus consensuelle, c'est-à-dire, donnant le moins lieu à des effets discriminants, n'est mentionnée dans aucune des deux méthodes précédentes prises séparément : il s'agit de la classe *travail-entreprise*.

The first congress of the French Sociological Association was an occasion to provide a synthesis of the subjects of conversation current in French sociology, and we will attempt to analyze this vocabulary and its frequencies. This article confronts this first analysis with a more quantified one, based on the distance from the means, to find how the two approaches can be combined. The lexical analysis, with the Alceste program, stresses the co-occurrence frequencies when words appear together in an abstract. The most frequent co-occurrences are found in the class "society-inequality", and the class "theory-authors". Calculating the percentage difference from the means (program Trideux) stresses the differences, and puts into comparable position similar differences. In this case, the words that differ the least from the means will be those used the most by sociologists and are found in the classes "method" and "public polity". If the authors are introduced into the analysis with variables for sex, institutional origin, location, and are associated with the words they use, we can combine the two methods and still conserve the families of words and those words which are the most significant, while cross-tabulating them

with accumulated individual characteristics. The most discriminating variable est thus sex, with a concentration of women in the family "society-inequality", and a concentration of men in the family "theory-authors". The class on which there is the most agreement between the two methods is not mentioned individually in either method. It is the class "work-company".

INDEX

Keywords : Vocabulary, Occurrence of Key Words, Alceste, Trideux, Complementarity of Methods, French Sociological Association Congress, France, Sociology

Mots-clés : Lexique, Occurrences de mots-clés, Complémentarité des méthodes, Congrès de l'AFS, Sociologie

AUTEUR

GÉRARD BOUDESSEUL

IUFM et Université de Caen Basse-Normandie ; gerard.boudesseul@caen.iufm.fr